Journal of Experimental Psychology:
Human Perception and Performance
1984, Vol. 10, No. 3, 340-357

Are Lexical Decisions a Good Measure of Lexical Access? The Role of Word Frequency in the Neglected Decision Stage

David A. Balota and James I. Chumbley
University of Massachusetts—Amherst

Three experiments investigated the impact of five lexical variables (instance dominance, category dominance, word frequency, word length in letters, and word length in syllables) on performance in three different tasks involving word recognition: category verification, lexical decision, and pronunciation. Although the same set of words was used in each task, the relationship of the lexical variables to reaction time varied significantly with the task within which the words were embedded. In particular, the effect of word frequency was minimal in the category verification task, whereas it was significantly larger in the pronunciation task and significantly larger yet in the lexical decision task. It is argued that decision processes having little to do with lexical access accentuate the word-frequency effect in the lexical decision task and that results from this task have questionable value in testing the assumption that word frequency orders the lexicon, thereby affecting time to access the mental lexicon. A simple two-stage model is outlined to account for the role of word frequency and other variables in lexical decision. The model is applied to the results of the reported experiments and some of the most important findings in other studies of lexical decision and pronunciation.

Psychologists have long been interested in the processes involved in word recognition. In studying variables that affect the speed of lexical access, researchers have relied heavily upon the lexical decision task (LDT). In this task the subject simply determines whether a letter string is or is not a word. When the manipulation of a variable causes a corresponding variation in response latency in the LDT, it has usually been assumed that the variable is having an effect on the ease of extracting sufficient information from a letter string to recognize it as a word, that is, to access its lexical representation. Obviously, this research technique makes the crucial assumption that lexical access is the only process in the LDT being affected by the manipulated variable. The results of experiments presented in this article have led us to question the validity of this assumption with respect to one important variable, word frequency. Specifically, we argue that the demand characteristics of the decision process in the LDT may result in an exaggerated role of word frequency. Finally, we propose a framework incorporating task-specific decision processes, and we use it to interpret data from the LDT. This framework aids in understanding how word frequency produces its effect in a number of different experimental situations.

The present research was initiated when an experimental result was encountered that was in conflict with the role of word frequency in three currently dominant models of word recognition: Morton's (1969, 1970, 1982) classic logogen model; Becker's (1976, 1979, 1980; Becker & Killion, 1977) verification model; and Forster's (1976, 1979) "bin" model. Each of these models places major emphasis on the role of word frequency in lexical access. The basic finding addressed by these models is that high-frequency words are recognized more quickly than low-frequency words. Full exposition of these models is provided in the

This research was partially supported by a National Institute of Mental Health postdoctoral fellowship (5-25007) to the first author and a research grant from the Centronics Corporation to the second author. Both authors contributed equally to this research project.

We acknowledge the able assistance of Hayley Arnett throughout this study. We also thank Charles Clifton, Jr., Janet Duchek, Jerome Myers, James Neely, Alexander Pollatsek, and Keith Rayner for comments on an earlier version of the manuscript. David Balota is now at the University of Kentucky.

Requests for reprints should be sent to James I. Chumbley, Department of Psychology, University of Massachusetts, Amherst, Massachusetts 01003.

earlier references, but there are three important assumptions common to all three models that are relevant to the present research: (a) Lexical access involves some matching of the features extracted from the stimulus to an internal representation of words; (b) word frequency determines the availability of lexical representations either by ordering them or by affecting their thresholds; (c) higher order semantic information for a word presented in isolation becomes available only after lexical access has taken place. These assumptions are important because the present research raises serious questions about whether the results from the LDT, the principal task used to investigate lexical access, can be unequivocally used as support for the assumptions.

The three assumptions common to these models are intended as a general characterization of the process of lexical access and, as such, should not be task dependent in their applicability. All tasks that involve lexical access should reflect these basic assumptions. Recently, we unexpectedly found that there was virtually no effect of word frequency in a task that should involve lexical access. The task was a simple category-exemplar verification task in which a category name (e.g., bird) was first presented and then was followed 800 ms later by an exemplar from that category (e.g., robin) or from a different category (e.g., sofa). The subject's task was to make a *yes–no* judgment about the validity of the category-exemplar relationship being presented. Although the results of this study yielded an interesting pattern of effects of instance dominance (likelihood of producing the exemplar given the category name) and category dominance (availability of the category name given the exemplar) on verification time, there was virtually no influence of word frequency on either trials in which a *yes* response was correct or, more important, on trials in which a *no* response was correct. The data from the *yes* trials are of marginal interest because one might expect a reduced effect of word frequency on these trials since the category name may have semantically primed the exemplars of that category, and this priming may have diminished any frequency effect. In fact, Becker (1979) has recently reported such an interaction between prime relatedness and word frequency in an LDT.[1] On the other hand, the lack of a word-frequency effect in the category-exemplar *no* response data is considerably more difficult to dismiss. On these trials, the category name could not have semantically primed an exemplar from another category. Therefore, the subject had to access the lexiçal entry for the exemplar and only then extract sufficient semantic information to make a decision. Because word frequency presumably influences the lexical access process, it clearly should have had an effect on the trials in which a *no* response was required.

In pursuing this theoretically discrepant finding, we decided to replicate the category-verification study and increase the power (by doubling the number of subjects) to detect a word-frequency effect. The approach taken was to simultaneously consider a number of variables, some that should influence lexical access and others that should not. Therefore, word frequency and the length of the word in letters and in syllables, (variables from the former class) were considered along with Battig and Montague's (1969) instance dominance and a measure of category dominance based on category selection time. Because we were interested in the unique effect of word frequency above and beyond the influence of these other variables, we utilized a multiple regression analysis. Through this approach, we could use a relatively large set of words without restricting our selection of items to those that orthogonally vary on each of the variables of interest. Attempts to orthogonally manipulate several lexical and semantic variables simultaneously could lead to a set of stimuli containing unusual items that are unrepresentative of the general population of words. Consider, for example, the difficulties

[1] It is worth noting here that even on a category-exemplar *yes* trial, one should expect some, albeit reduced, frequency effect. That is, in the Chumbley (1984) study half of the words had high instance dominance ratings and half had low instance dominance ratings, as measured by Battig and Montague (1969). Because it seems unlikely that a category name will sufficiently prime the items with low instance dominance (e.g., *furniture–mirror*) to completely override the word-frequency effect, one should simply find a reduced effect. This is especially the case when one considers that in the Becker (1979) study a 49-ms word-frequency effect was still found in his LDT for the related targets that followed highly associated primes.

Table 1
Intercorrelations, Means, and Standard Deviations for the Predictor Variables

Predictor variable	1	2	3	4	5
1. IDOM	1.00				
2. LFREQ	.45	1.00			
3. CDOM	.33	.05	1.00		
4. LENG′	.30	.46	.12	1.00	
5. SYLL′	.21	.36	.16	.70	1.00
M	144.61	51.27	−962.22	−5.74	−1.65
SD	153.07	6.76	177.72	1.57	0.63
Tolerance	.68	.68	.82	.45	.50

Note. IDOM = instance dominance; LFREQ = log word frequency; CDOM = category dominance; LENG′ = reflected number of letters; SYLL′ = reflected number of syllables. The CDOM, LENG′, and SYLL′ are reflected values obtained by multiplying by −1. These reflections were performed to place these predictors in the same relationship to RT as IDOM and LFREQ, that is, the larger the value of the lexical variable, the shorter the expected RT. Tolerance = 1 − the square of the multiple correlation of one predictor variable with the remaining predictor variables.

one would encounter in finding a high-frequency word that has low category dominance, high instance dominance, nine letters, and only one syllable. Faced with such a trade-off, we elected to use multiple regression techniques to examine the role of these variables.

Experiment 1

Method

Subjects. Twenty undergraduate students recruited from the subject pool at the University of Massachusetts, Amherst, participated in partial fulfillment of course requirements. No subject participated in more than one of the present experiments.

Apparatus. The experiment was controlled by a North Star Horizon computer. Stimulus words were displayed in uppercase letters on a television monitor driven by an IMSAI memory-mapped video raster generator. In order to increase legibility, stimuli were presented with a single space between letters. Subjects were seated approximately 50 cm from the video monitor. A three-letter word (three letters separated by two spaces) occupied a visual angle of approximately 1.1°, whereas a nine-letter word occupied a visual angle of approximately 3.7°. Reaction time (RT) and interval timing were both measured with millisecond (ms) accuracy via the computer. The same apparatus was also used in Experiments 2 and 3.

Materials. A total of 72 target words was selected from Rosch (1975) for use in this study. They consisted of eight exemplars from each of nine different categories. Each category was represented by four high-typical and four low-typical exemplars. (The complete list of target words along with each word's mean reaction times and percentage of error rates for all the experiments is available from the authors.) A total of 50 buffer/practice words was selected with 5 instances from each of 10 categories in the Battig and Montague (1969) norms. These buffer/practice words approximately matched the targets in both length and syllables and were not members of any of the target categories.

The category dominance measure was obtained from an independent group of 20 subjects from the same subject pool. Subjects were first given a list of the nine category names to memorize. They were then given repeated trials on which an exemplar of one of the nine categories was presented, and the task was to say aloud the name of the category to which the exemplar belonged. The sound of the subject's saying the category name triggered a voice key. The vocal RT is the category selection time (Sanford & Seymour, 1974) for the exemplar. Although full details will be reported in a future report (Chumbley, 1984), it is noteworthy that subjects were well practiced with the category names, and each exemplar was tested several times.

Table 1 presents the correlation matrix, mean values, and standard deviation of these values for each of the lexical variables used in the studies reported here. Each measure is based on the values for the 72 target words. Instance dominance (IDOM) was measured by using the Battig and Montague (1969) norms. Based on observations made by Whaley (1978), log word frequency (LFREQ) was used instead of raw frequency (*f*), determined from the Kučera and Francis (1967) norms, and the following function: LFREQ = 40 + 10 log(*f* + 1). Category dominance (CDOM) is simply the category selection time for each word multiplied by −1.

The tolerance values given in Table 1 are measures of the intercorrelation of the predictor variables. They represent the extent to which a predictor variable is simply a linear combination of the other predictor variables entered into the regression. The maximum tolerance value possible is 1.0 (totally orthogonal predictor), and the minimum tolerance value is 0 (totally predicted by the other variables). As can be seen, each predictor variable, with the exceptions of length and syllables, has a relatively high tolerance with the other variables. The tolerance values of length and syllables are reduced primarily because of their high (.70) intercorrelation.

Procedure. None of the subjects knew in advance which categories would be used in the experiment. Each subject was presented two blocks of 30 practice trials before being tested with the target words. Thus, subjects experienced 60 practice trials using the 10 buffer/practice nontarget categories before responding to words from any of the

target categories. Following the two blocks of practice trials, eight blocks of test trials were presented. Each block had 36 test trials that followed 5 buffer trials at the beginning of each block. The eight test blocks were divided into two cycles. Each of the 72 target words was presented twice in a cycle, once with the appropriate category name and once with one of the other eight target category names. These 144 conditions were randomly ordered within each cycle.

On each trial the following sequence of events occurred: (a) a 500-*Hz* warning tone presented for 250 ms; (b) a 250 ms-interstimulus interval; (c) a category name presented for 800 ms; (d) an exemplar presented below the category name until the subject responded by pulling one of two levers; (e) either a 2-s intertrial interval or an error message that was terminated by the subject's pressing a third button (placed between the response levers) that was then followed by the 2-s interval before the start of the next trial.

All subjects were tested individually in a sound-deadened room. Two response levers, one designated *yes* and the other *no,* were placed so that the subject's fingers could comfortably rest against them. Subjects were instructed to decide as quickly as possible whether each exemplar was a member of the designated category while maintaining an accuracy level of at least 95% correct. Feedback regarding both RT and accuracy was given after each block of trials. There was a 10-s mandatory rest between blocks followed by a signal that the subject could continue the experiment by a button press when ready. Ten subjects used their dominant hand to respond *yes,* and 10 used their nondominant hand to respond *yes.*

Results

There were four sets of data for each subject defined by the factorial combination of Response (*yes* or *no*) × Cycle (Test Blocks 1–4 or Test Blocks 5–8). Each subject's data for each condition were scored for correctness and for outliers in RT. Outliers were defined as being either (a) less than 200 ms or (b) longer than 2 s and also more than three standard deviations above the subject's mean for the condition. Errors and outliers were replaced with the subject's mean correct response RT for that condition. The mean percentage of error rate per subject (across conditions) was 4.48, and the mean percentage of outlier rate was 0.12.

A full multiple regression analysis was conducted on the data for each response and each cycle. This procedure parallels that used by other workers investigating category verification (Anderson & Reder, 1974; Loftus & Suppes, 1972) and lexical decision (Whaley, 1978). The full analysis was used instead of a stepwise analysis because we had identified in advance which variables were of theoretical importance, and we were not primarily interested in comparing the relative importance of the variables within a task. Our concern was in comparing the relative importance of a given variable in one task to its importance in other tasks. The mean RT across subjects for each word was determined, and these means were the criterion variable in the regression analysis. The five predictor variables were IDOM, LFREQ, CDOM, LENG′ (length in letters), and SYLL′ (number of syllables).

When predictor variables have widely different ranges, as is the case here, the actual value of the regression coefficient (*Beta*) is not very helpful in evaluating the size of the effect produced by a predictor variable. For this reason, we have chosen to present what will be referred to as a *semistandardized regression coefficient.* These coefficients are simply the product of *Beta′* and the standard deviation of the criterion variable. The correlations, *F*s, and partial correlations are unchanged by this procedure. The semistandardized regression coefficient indicates the change in RT with one standard deviation unit change in the predictor variable. Table 2 presents the semistandardized regression coefficients and the raw and partial correlations with RT for each predictor variable.

Through the use of the semistandardized regression coefficients, one can compare the size of the effects of the predictor variables in this experiment with those in other studies. Approximately 95% of the words had predictor variable values within two standard deviation units of either side of the mean. Multiplying the semistandardized regression coefficient by 4 yields the approximate change in RT produced by changing the value of a predictor variable from its lowest value to its highest value. Thus, LFREQ produced only a small 24-ms effect for *yes* responses in Cycle 1, whereas IDOM produced a large 160-ms effect, and CDOM produced an even larger 200-ms effect. It should be noted that these are unique effects over and above the effects jointly produced with the other variables.

The results of the multiple regression analyses indicated that only IDOM, CDOM, and LENG′ had any significant unique relationship to RT in category verification. For the *yes* responses, the regression coefficients for IDOM were highly significant in both Cycle 1, $F(1,$

66) = 19.72, $p < .001$, and Cycle 2, $F(1, 66) = 14.45$, $p < .001$. Similarly, the CDOM regression coefficients were highly significant for *yes* responses in both Cycle 1 and Cycle 2, $F(1, 66) = 47.01$, $p < .001$, and $F(1, 66) = 37.03$, $p < .001$, respectively. For *no* responses, LENG' in Cycle 1, $F(1, 66) = 4.53$, $p < .05$, and IDOM in Cycle 2, $F(1, 66) = 4.76$, $p < .05$, had significant effects. The unique influence of LFREQ did not approach significance for either the *yes* responses (both Fs < 1.10) or for the *no* responses (both Fs < 2.24). None of the remaining predictor variables yielded significant regression coefficients. The value of R^2 for each of the regression analyses was significant, all Fs(5, 66) > 2.85, ps < .05. These values were .66 for *yes* responses, Cycle 1; .59 for *yes* responses, Cycle 2; .22 for *no* responses, Cycle 1; and .18 for *no* responses, Cycle 2.

Each exemplar was actually presented twice during each cycle, once when a *yes* response was appropriate and once when a *no* response was appropriate. We were concerned that this might influence the size of the predictor variable effects, especially that of LFREQ. Regression analyses were conducted for the first presentation of an exemplar in either the *yes* or the *no* correct response condition of Cycle 1. The results of these analyses were quite clear. The unique effect of LFREQ was not significant on the very first presentation of an item as either a *yes*, $F(1, 66) < 1$, or a *no*, $F(1, 66) = 1.98$. This lack of a significant unique effect of LFREQ cannot be attributed to a lack of power because the pattern of significant effects was exactly the same as that displayed in Table 2 for both *yes* and *no* responses. Thus, it seems fair to conclude that word repetition within a cycle is not diluting the LFREQ effect.

Discussion

The results from Experiment 1 were quite straightforward. For the category-exemplar *yes* trials, there were large unique effects of IDOM and CDOM but not LFREQ. On the other hand,

Table 2
Regression Coefficients, Raw Correlations, and Partial Correlations for the Category Verification Experiment

Cycle	Predictor variable				
	IDOM	LFREQ	CDOM	LENG'	SYLL'
	Yes responses				
1					
Regression coefficient	−37.62	−6.07	−52.62	−16.45	6.09
r	−.64	−.32	−.66	−.21	−.07
Partial r^2	.23	.01	.42	.04	.01
2					
Regression coefficient	−25.95	−7.11	−37.64	−4.71	2.25
r	−.60	−.31	−.65	−.14	−.04
Partial r^2	.18	.02	.36	.00	.00
	No responses				
1					
Regression coefficient	−4.81	−10.53	−7.77	−18.36	7.29
r	−.30	−.36	−.17	−.35	−.17
Partial r^2	.01	.03	.02	.06	.01
2					
Regression coefficient	−9.42	−2.95	−2.67	−0.09	−3.58
r	−.39	−.27	−.17	−.20	−.20
Partial r^2	.07	.01	.01	.00	.01

Note. IDOM = instance dominance; LFREQ = log word frequency; CDOM = category dominance; LENG' = reflected number of letters; SYLL' = reflected number of syllables. The mean RT for *yes* responses, Cycle 1, was 662.96 ms (SD = 96.82 ms; for *yes* responses, Cycle 2, it was 606.61 ms (SD = 71.56 ms); for *no* responses, Cycle 1, it was 693.33 ms (SD = 53.29 ms); and for *no* responses, Cycle 2, it was 628.63 ms (SD = 31.80 ms).

for the category-exemplar *no* trials there were only small effects of LENG′ during Cycle 1 and IDOM during Cycle 2. These results suggest that in verifying that an exemplar is a member of a category, the subject may access the category name (as evidenced by the strong effect of CDOM on the *yes* responses). For the present purposes, however, the more important finding is that there was little, if any, unique impact of word frequency on either the category verification *yes* or *no* trials. As noted earlier, the lack of a robust frequency effect for the *no* trials is particularly perplexing because it is unclear, within the currently available models of lexical access, why word frequency should not produce an independent influence on a misprimed category-verification *no* decision (e.g., bird–sofa), a decision that must certainly require lexical access. In this light, it is noteworthy that Anderson and Reder (1974) and Millward, Rice, and Corbett (1975) have also reported failures to find an impact of instance frequency on instance–category and category–instance *no* decision trials, respectively.

Before considering the implications of these results in more detail, a number of alternative explanations must be considered. One alternative explanation could be that there was only a very small frequency effect in the present category-verification task because there were multiple exemplars from the nine target categories, and some sort of implicit semantic priming could account for the lack of a unique frequency effect. Five converging lines of argument reduce the plausibility of this explanation. First, it is unclear how such priming could be effective on *no* response trials when the exemplar is being inappropriately primed by the name of another category. Second, there was little evidence that the frequency effect was disappearing across trials as subjects became more and more familiar with the categories. Third, prior to the first cycle, subjects should have actually been misprimed by the 10 nontarget buffer/practice categories because they received 65 buffer/practice trials using exemplars from these categories before they received any exemplars from the target categories. These items were also used in the 5 buffer trials at the beginning of each test block. Fourth, it will be seen in Experiment 3 that when the number of exemplars from each category was doubled, there was virtually no impact on the obtained frequency effect. This suggests that increasing the emphasis on the target categories did not appreciably influence the word-frequency effect, at least not in a pronunciation task. Finally, Becker (1979) found a significant frequency effect (49 ms) in a lexical decision task, albeit reduced in comparison to an unrelated priming condition, when each word was primed on a given trial by a high associate. It is unlikely that implicit priming for misdirected category-exemplar *no* trials could produce more semantic priming than that produced by highly related associates.

Although the implicit priming account of the results of the first experiment is inadequate, there are two simpler accounts that must be addressed. First, the range of word frequency for the words utilized in Experiment 1 may not have been sufficiently large to produce a robust frequency effect. The words were originally selected so that college sophomores would have little difficulty knowing the meaning of even the lowest frequency words used in the current studies. It is possible that part of the *word-frequency effect* reported in the literature can be attributed to subjects' not knowing the meaning of a particular word rather than the frequency of occurrence of that word in print. For example, in one study investigating word-frequency effects (Frederiksen & Kroll, 1976) error rates in excess of 40% were reported for the low-frequency targets in an LDT where chance performance is 50%.

A second possibility is that the word-frequency effect found in previous studies was actually an effect of other variables, such as, IDOM and LENG′, which covary with word frequency. Because our major interest was in the unique effect that could be unequivocally attributed to word frequency, we used the regression analysis to remove the joint effects of other variables. It is possible, therefore, that there is very little unique effect of word frequency on lexical access above and beyond these covarying variables.

The most obvious way to test these possibilities is to conduct a lexical decision experiment with the same set of words and the same set of predictor variables. This was accomplished in Experiment 2. If the lack of a large unique frequency effect in Experiment 1 was

Table 3
Regression Coefficients, Raw Correlations, and Partial Correlations for Word Responses in the Lexical Decision Experiment

Cycle	Predictor variable				
	IDOM	LFREQ	CDOM	LENG′	SYLL′
1					
Regression coefficient	−10.29	−24.29	−10.66	−19.46	6.09
r	−.53	−.66	−.25	−.51	−.31
Partial r^2	.05	.24	.07	.12	.01
2					
Regression coefficient	−10.01	−19.59	−7.76	−11.99	4.71
r	−.54	−.63	−.26	−.44	−.26
Partial r^2	.07	.22	.05	.06	.01

Note. IDOM = instance dominance; LFREQ = log word frequency; CDOM = category dominance; LENG′ = reflected number of letters; SYLL′ = reflected number of syllables. The mean RT for Cycle 1 responses was 570.82 ms (*SD* = 55.04 ms), and for Cycle 2 responses it was 543.22 ms (*SD* = 44.77 ms).

simply due to the fact that the words or analysis procedures somehow did not allow a frequency effect to be demonstrated, then frequency should have little impact in the following lexical decision experiment.

Experiment 2

Method

Subjects. Twenty undergraduate students were recruited from the same pool described in Experiment 1.

Materials. The words in Experiment 1 were used as targets in Experiment 2. An additional 122 words were chosen from the Battig and Montague (1969) norms. These words were from the same categories as the target items (8 from each of the 9 target categories and 5 from each of the 10 buffer/practice categories). Pronounceable nonwords were produced by changing up to three letters within these words (e.g., *fishing* was changed to *fisleng*). The mean length of the nonwords (5.76) closely matched the mean length of the target words (5.74).

Procedure. The procedure in Experiment 2 was the same as that in Experiment 1 except that now the subject's task was to make a lexical decision about the exemplar or nonword, and these stimuli were not preceded by the category name. Subjects indicated their decision by pulling one of the two response levers. All target words and their nonword counterparts occurred once during the Test Blocks 1–4 (Cycle 1) and once during the Test Blocks 5–8 (Cycle 2). Prior to the presentation of the test blocks, subjects were presented two practice blocks using the practice/buffer items. The subjects were not informed about the categorical structure of the stimuli. Across subjects, dominant and nondominant hands were balanced across *word* and *nonword* responses.

Results

Regression analyses, as described above, were performed on word RT (corrected for errors and outliers and then averaged across subjects as in Experiment 1) with the five predictor variables previously used. The results of these analyses for Cycles 1 and 2 may be seen in Table 3. The mean percentage of error rate was 3.89, and the mean percentage of outlier rate was 1.25.

The results of Experiment 2 are strikingly different from those of Experiment 1. The regression coefficients for LFREQ are very large both in absolute terms and in comparison to those found in category verification. LENG′ had a fairly large effect in Cycle 1 but had a reduced effect in Cycle 2. CDOM and IDOM produced relatively small regression coefficients compared to their large effects on *yes* RT in Experiment 1. Multiplying the semistandardized regression coefficients by 4 produces approximately a 100-ms unique effect of LFREQ in lexical decision, Cycle 1, but only 40-ms effects for IDOM and CDOM.

The results of the multiple regression analyses supported the above observations. The LFREQ regression coefficient was highly significant in both Cycle 1, $F(1, 66) = 21.05$, $p < .001$, and Cycle 2, $F(1, 66) = 18.31$, $p < .001$. The regression coefficient for LENG′ was significant in both Cycle 1, $F(1, 66) = 9.00$, $p < .01$, and Cycle 2, $F(1, 66) = 4.57$, $p < .05$. Of the remaining variables, only the regression coefficients for CDOM in Cycle 1 and IDOM in Cycle 2 were significant, $F(1, 66) = 4.92$, $p < .05$, and $F(1, 66) = 4.75$, $p < .05$, respectively. The values of R^2 were highly significant for both Cycle 1 and Cycle 2, both Fs(5, 66) >

14.98, $p < .01$. R^2 for Cycle 1 was .59 and for Cycle 2 it was .53.

Discussion

The results of the second experiment were again quite clear and differed dramatically from the results of the first experiment. The same set of words and predictor variables were used in both experiments. Results of the second experiment yielded large effects of LFREQ over and above the effects of the other variables, whereas Experiment 1 yielded little evidence of a unique frequency effect in category verification. These contrasting results are inconsistent with explanations of the results of Experiment 1 that attribute the small word-frequency effect to properties of the words or of the regression analysis technique.

Before we describe a possible account of the differing results of the first and second experiments, there is another noteworthy result of the second experiment. It was found that both CDOM and IDOM significantly predicted lexical decision performance over and above their shared predictiveness with LFREQ, LENG′, and SYLL′. This finding is important because it has typically been argued that semantic information becomes available only after lexical access. In this light, the results of the second experiment indicate that either this argument is incorrect or that some other component of lexical decision is sensitive to meaning variables. Other investigators have reported similar findings. James (1975), Whaley (1978), and Chumbley and Balota (1984) have reported effects of semantic variables in lexical decision performance. Thus, there is mounting evidence that the LDT is not a good tool for studying a lexical access process that is presumed to be unaffected by semantic variables.

Recently, a number of theorists (Forster, 1979; Theios & Muise, 1977; West & Stanovich, 1982) have suggested that the LDT involves postrecognition processing that may influence performance. These theorists further suggest that when one considers contextual effects, the pronunciation task may be a better reflection of pure lexical access (see, however, Coltheart, Davelaar, Jonasson, & Besner, 1977). In light of this possibility, it was decided to use the set of materials and predictor variables from the previous two experiments in a pronunciation task. Very simply, if the semantic effects found in the second experiment reflected postaccess semantic influences on the decision process, then one might expect these effects to be eliminated in a pronunciation task.

In Experiment 3 we also attempted to address whether the semantic effects in the second experiment may have been due to a type of implicit priming. That is, across trials, the categories and their respective names may have become sufficiently activated so that subjects were using semantic information in making their lexical decisions. A second possibility is that, across trials, items from the same semantic category may have occurred on adjacent trials, thereby producing some intertrial semantic priming. In an attempt to address these possibilities, and because we were not interested in the pronunciation of nonword items, we replaced the nonwords used in the second experiment with either the word *blank* or eight other exemplars from each of the nine semantic categories. If the semantic effects found in the LDT were simply due to the loading of the semantic categories, then we should find larger semantic effects in the condition where there are twice as many exemplars from each of the categories.

Experiment 3

Method

Subjects. Forty undergraduate students were recruited from the same subject pool described earlier. Twenty subjects participated in the word/word condition and 20 in the word/blank condition.

Materials. The words from the earlier experiments were again used in Experiment 3. For the word/word condition, the nonwords used in Experiment 2 were replaced by the words from which those nonwords were derived (e.g., *fisleng* was replaced by *fishing*). For the word/blank condition, the word *blank* was simply presented on half the trials. Thus, on test trials in which a nonword was presented during Experiment 2, either a word from one of the target categories was presented or the word *blank* was presented.

Procedure. In this experiment, subjects were simply asked to pronounce each word aloud. A voicekey connected to the computer detected onset of their pronunciation and measured latencies to the nearest millisecond. When a pronunciation was detected, the message "Response OK?" immediately replaced the word. Subjects were instructed to pull the right lever if they felt that their correct pronunciation of the word triggered the computer and to pull the left lever if they incorrectly pronounced the word or if some other auditory sound (such as a cough) triggered

the computer. If the OK lever was pulled, there was a 2-s interval before the warning tone was presented to begin the next trial. If the error lever was pulled, the subject had to press the third button when ready to begin the 2-s intertrial interval.

Results

Regression analyses were performed as before on the data for each cycle of each group. The mean percentage of error rates for the word/word and word/blank groups were 1.46 and 1.42, respectively, whereas the corresponding mean percentage of outlier rates were 0.8 and 1.01.

The regression coefficients for LFREQ seen in Table 4 appear to be smaller than those found for lexical decision (see Table 3). It seems there is a smaller unique frequency effect in the pronunciation tasks (about 50 ms) than in the LDT (about 100 ms). Andrews (1982) and Frederiksen and Kroll (1976) have found similar differences between the LDT and the pronunciation task. The unique effect of LFREQ on pronunciation, however, still is larger than its 24-ms effect on category verification. LENG′ and CDOM have about the same effect in lexical decision and pronunciation. The regression coefficients for IDOM are noticeably smaller for the pronunciation tasks than for either the LDT or category verification task.

The results of the multiple regression analyses indicated that the regression coefficient for LFREQ was highly significant in the word/word conditions for Cycle 1, $F(1, 66) = 14.86$, $p < .001$, and Cycle 2, $F(1, 66) = 14.23$, $p < .001$, and the word/blank conditions for Cycle 1, $F(1, 66) = 13.27$, $p < .001$, and Cycle 2, $F(1, 66) = 10.27$, $p < .01$. Similarly, the LENG′ regression coefficients were highly significant in the word/word conditions for Cycle 1, $F(1, 66) = 21.64$, $p < .001$, and for Cycle 2, $F(1, 66) = 18.19$, $p < .001$, and in the word/blank condition for both Cycle 1, $F(1, 66) = 6.08$, $p < .05$, and Cycle 2, $F(1, 66) = 23.68$, $p < .001$. The only remaining coefficients to reach significance in the multiple regression analyses were those for CDOM for the word/word condition, Cycle 2, $F(1, 66) = 8.47$, $p < .01$, and

Table 4
Regression Coefficients, Raw Correlations, and Partial Correlations for the Pronunciation Experiment

Cycle	Predictor variable				
	IDOM	LFREQ	CDOM	LENG′	SYLL′
Word/word condition					
1					
Regression coefficient	−4.36	−16.49	−6.78	−24.84	4.33
r	−.45	−.63	−.15	−.65	−.41
Partial r^2	.01	.18	.04	.25	.01
2					
Regression coefficient	2.11	−11.45	−8.01	−15.86	−2.51
r	−.34	−.60	−.17	−.66	−.50
Partial r^2	.01	.18	.11	.22	.01
Word/blank condition					
1					
Regression coefficient	−2.55	−15.66	−6.16	−12.99	−4.75
r	−.40	−.60	−.13	−.57	−.46
Partial r^2	.01	.17	.04	.08	.01
2					
Regression coefficient	2.38	−10.23	−7.54	−19.04	3.28
r	−.31	−.54	−.17	−.63	−.38
Partial r^2	.01	.13	.09	.26	.01

Note. IDOM = instance dominance; LFREQ = log word frequency; CDOM = category dominance; LENG′ = reflected number of letters; SYLL′ = reflected number of syllables. The mean RT for the word/word condition, Cycle 1 was 492.05 ms (*SD* = 46.21 ms); for word/word, Cycle 2, it was 457.08 ms (*SD* = 32.09 ms); for word/blank, Cycle 1, it was 485.66 ms (*SD* = 41.08 ms); and for word/blank, Cycle 1, it was 461.08 ms (*SD* = 31.42 ms).

for the word/blank condition, Cycle 2, $F(1, 66) = 6.76$, $p < .05$. The R^2 values for each condition were highly significant, all Fs(5, 66) > 13.74, all ps < .001. The values were .60 for word/word, Cycles 1 and 2; .51 for word/blank, Cycle 1; and .54 for word/blank, Cycle 2.

Discussion

The results of the third experiment indicated that the pattern of data for the word/word and the word/blank conditions was virtually identical. That is, in both conditions the length of the word and its frequency were strong predictors of pronunciation latencies over and above their joint effect with the other variables. Furthermore, the only impact of a semantic variable was the effect of CDOM during Cycle 2 and, interestingly, this factor had approximately the same size effect for both the word/word and word/blank conditions. Thus, the semantic effects being produced in Experiments 2 and 3 do not appear to be due simply to exemplars from the same semantic category being presented.

One possible reason that CDOM predicted pronunciation latency is that both the category selection and the pronunciation tasks involve the same lexical access process, that is, one has to recognize a word to produce its category name. Of course, this could also account for the CDOM effect found in the results of the lexical decision experiment. In the lexical decision experiment, however, IDOM also significantly predicted performance. In this light it is noteworthy that IDOM did not predict pronunciation performance in either Cycle 1 or Cycle 2 (both Fs < 1). Thus, the IDOM effect found in the LDT may reflect a postaccess influence that is not reflected in pronunciation performance, consistent with the recent views regarding the LDT cited in the introduction to Experiment 3.

It is also important that the LFREQ effect was the same in the word/word and word/blank conditions. As noted earlier, if implicit semantic priming is the reason for the absence of a LFREQ effect in Experiment 1, there should have been a decrease in the LFREQ effect in the word/word conditions because each semantic category was represented by twice as many words as in the word/blank condition.

In the discussions above, comparisons of the sizes of the regression coefficients across tasks were made without reference to the statistical reliability of these differences. We now present statistical confirmation of these differences.

Overall Analysis Section

The analysis procedure presented here adopts a different perspective in analyzing the data. A regression analysis was conducted on the data for each subject. The average regression coefficient for each predictor variable for a given task was then computed, and the variability of the regression coefficient across subjects within a task was used to test whether different tasks requiring lexical access involved similar effects of word frequency. The appropriate test simply involved conducting an analysis of variance on the regression coefficients for subjects performing different tasks. Because each experiment used the same set of words and the same set of predictor variables, the intercorrelations among the variables was a constant, and the task was the equivalent of a variable manipulated orthogonally to these predictor variables. Therefore, any change in a regression coefficient from one task to another must be due to a change in the way the predictor variable is related to RT in the tasks.

Because response was a repeated measures factor in the category verification task, two sets of analyses of variance were performed. One set incorporated the individual regression coefficients from the *yes* response condition, and the other used the *no* response data. Each set of analyses included an analysis for each of the five predictor variables. The results of the analyses for SYLL′ will not be reported because, as in all of the analyses reported earlier, it did not produce any significant unique effects. The mean regression coefficients from these eight analyses are presented in Table 5. The means presented are averaged across cycles because none of the interactions of Task × Cycle reached significance, all Fs(3, 76) < 2.30.

The most important finding displayed in Table 5 is that the effect of LFREQ varied significantly across the different tasks with $F(3, 76) = 4.97$, $p < .01$, standard error of the mean (SE_M) = 2.82, for the *yes* analysis, and

Table 5
Mean Regression Coefficients Averaged Across Cycles From the Individual Subject Regression Analyses

Condition	Predictor variable			
	IDOM	LFREQ	CDOM	LENG′
Lexical decision	−10.15	−21.94	−9.22	−15.72
Pronunciation word/word	−1.12	−13.97	−7.39	−20.35
Pronunciation word/blank	−0.08	−12.94	−6.85	−16.01
Category verification *yes*	−31.78	−6.59	−45.13	−10.58
Category verification *no*	−7.12	−6.74	−5.22	−9.22

Note. IDOM = instance dominance; LFREQ = log word frequency; CDOM = category dominance; LENG′ = reflected number of letters; SYLL′ = reflected number of syllables.

$F(3, 76) = 7.22$, $p < .01$, $SE_M = 2.32$, for the *no* analysis. As can be seen, LFREQ had a large effect on lexical decision, a moderate effect on pronunciation, and a very small effect on category verification. Three contrasts confirmed this observation. The average regression coefficient for lexical decision differed from that for pronunciation, $t(76) = 2.45$, $p < .02$, and the average regression coefficient for pronunciation differed from that for category verification for both *yes*, $t(76) = 1.99$, $p < .05$, and *no* responses, $t(76) = 2.36$, $p < .02$.

The analyses of variance indicated that IDOM also had a significant effect across tasks with $F(3, 76) = 44.00$, $p < .001$, $SE_M = 2.22$, for the *yes* analysis and $F(3, 76) = 3.46$, $p < .05$, $SE_M = 2.59$, for the *no* analysis. Contrasts performed on the regression coefficients for IDOM indicated that this variable had a greater effect for category verification *yes* responses than for either lexical decision or category verification *no* responses (nondirectional $ps < .05$), but the latter two tasks did not differ from each other. The effect of IDOM in pronunciation was less than that in any of the other tasks (all nondirectional $ps < .05$). The analysis for CDOM *yes* responses yielded a highly significant effect, $F(3, 76) = 41.68$, $p < .001$, $SE_M = 2.89$, whereas that for the *no* responses produced an $F < 1$, $SE_M = 2.42$. Contrasts confirmed what is fairly obvious: The coefficient for category verification *yes* responses significantly differed from all the other coefficients that did not differ from each other. The analyses with respect to LENG′ yielded no significant task effects with both $Fs(3, 76) < 1.58$ ($SE_M = 3.45$ for *yes* responses and 3.66 for *no* responses).

One final analysis was conducted to eliminate a possible concern of some readers. One might argue that the multiple regression analyses may have been somehow permitted variables relevant to the category verification task, but not to lexical decision, to mask the effect of frequency in category verification. For example, IDOM should be more related to the category verification task than to the LDT. Also, IDOM is correlated with LFREQ. Thus, according to such an argument, IDOM could have accounted for some of the variance in the category verification task that LFREQ would have otherwise predicted. Fortunately, there is a simple way (for our data) of addressing this possibility. We simply need to compare the raw correlation between LFREQ and category verification RT to that between LFREQ and lexical decision RT. Obviously, these raw correlations are not affected by competing variables. If these correlations are significantly different, then we can be sure that the regression analyses are not producing misleading results. A Fisher's Z′ test of the difference in these two raw correlations indicated that the −.30 correlation between LFREQ and category verification *no* response latency was significantly different from the −.64 correlation between LFREQ and LDT response latency, $p < .025$. Furthermore, the differences between the category verification *no* correlation and those for the pronunciation tasks were also significant (all nondirectional $ps < .054$). Thus, these analyses rule out the possibility that the present differences across tasks are simply due to the way in which shared variance is partialed in the regression analyses.

General Discussion

The most important finding in the present results is that word frequency was highly related to lexical decision performance over and above its joint relationship with other variables but had little unique relationship to category verification performance even though lexical access must be involved in the category ver-

ification task. Independent of what one chooses to believe about the relationship between word frequency and lexical access, these results clearly indicate that word frequency has dramatically different effects, depending upon the task that is used to assess lexical access. This conclusion has major importance because current conceptions of word recognition attribute a major role to the word frequency variable in determining how the mental lexicon is structured and accessed. Given this role in lexical access, word frequency should not have the large task-specific effects we have observed.

In this light, there are two other findings that are relevant here. First, LFREQ was virtually uncorrelated (−.05) with the CDOM variable in the present study. This finding is noteworthy, because subjects need to recognize the word to determine the category of a presented exemplar. Thus, we have yet another instance in which there is little effect of word frequency in a task that demands lexical access.

Second, there are two recent studies reported by Kliegl, Olson, and Davidson (1982, 1983) in which it was found that word frequency accounted for only about 1%–3% of the variance in fixation durations in reading when other potentially confounding variables (such as word length) were partialed out. These results are particularly disturbing because they suggest that the effect of word frequency on fixation duration in reading cannot be unequivocally attributed to an unconfounded index of frequency.[2]

Thus, the category verification task is not the only task that does not produce the large (100 ms) unique frequency effect found in the LDT. In the present study we have been unsuccessful in our attempts to explain away the large reduction in frequency effects in the category verification task. A different approach is to ask why there is such a large frequency effect in the LDT. Because lexical decision performance has been a primary source of data for the current models of lexical access, it is crucial that one understand all components relevant to the LDT, especially those components other than lexical access that may themselves be affected by variables viewed relevant to lexical access. A recent example of an analysis of the components of lexical decision has been provided by Morton (1982). He concludes that at least a portion of the word-frequency effect in lexical decision can be attributed to the operation of processes in the cognitive system rather than to logogen threshold differences produced by differences in word frequency. Although our analysis was developed independently of Morton's and assumes specific processes about which he may have reservations, the basic thrust of our proposals is in the same direction.

A Framework for Understanding the Lexical Decision Task

One possible reason that the LDT produces such large frequency effects is that the task places a premium on frequency information at the decision stage of the task. Obviously, the fact that lexical decisions are typically more than twice as long (500–600 ms) as normal reading rates (250 ms/word) suggests that there is much more involved in the LDT than simple lexical access.[3] Given this possibility, it seems

---

[2] There are a number of points to note here. Both Carpenter and Just (1983) and Mitchell and Green (1976) have reported an influence of word frequency on reading time measures. In Carpenter and Just's study, the effect was considerably smaller when length was controlled, and, furthermore, some of their materials contained low-frequency words for which subjects did not know the meaning. In Mitchell and Green's study the effect of length was only partially controlled for by considering the length of a three-word triad in which the target appeared instead of the length of the target word itself. Thus, it is unclear what the true impact of frequency (above and beyond other potentially confounding variables) was in these studies.

One final variable should be considered in these more natural reading situations. That is, it is possible that the contextual constraints for the high-frequency and low-frequency words may vary. For example, it seems that the word *water* is much more likely in the following context than the word *tonic*.

The cold glass of *water* quenched the man's thirst.
tonic

Such contextual variables are very difficult to tease apart from true word-frequency effects on lexical access.

[3] Obviously, one could argue that lexical decision latencies also involve response execution (pushing the button). In this same light, however, one must acknowledge the complexity of reading, that is, the reader must program eye movements and integrate the currently accessed word with ongoing comprehension. It does not seem reasonable that simple response execution can account for the large differences in latencies between reading and lexical decisions.

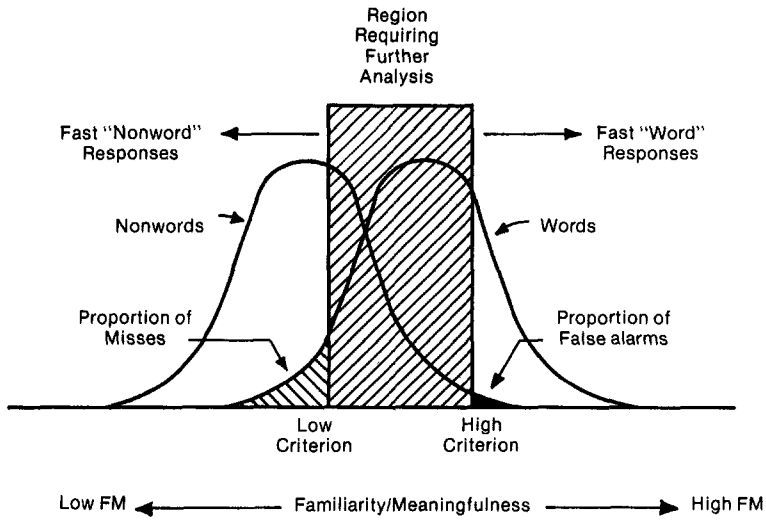

Figure 1. Word and nonword distributions along the familiarity/meaningfulness dimension.

necessary to consider more closely the task faced by the subject when making a lexical decision. Basically, the subject is asked to discriminate meaningful stimuli from nonword letter strings that, with the exceptions of misspellings, have never been seen before. The two most obvious pieces of information the subject could utilize to make such discriminations are the frequency with which the stimulus has been seen before and its meaningfulness. It is important to note here that the fact that frequency information is available does not provide evidence that this information in some way orders lexical access.

A variant of the two-stage model developed by Atkinson and Juola (1973) for the memory search task provides a framework that more closely considers the decision process in the LDT. This model is shown in Figure 1. The basic notion is that words and nonwords differ on a familiarity/meaningfulness (FM) dimension. A particular letter string's value on this FM dimension is based primarily on its orthographic and phonological similarity to actual words. The word and nonword distributions on the FM dimension are separated but overlap. The subject can use this fact in following the LDT instructions to both maximize speed and minimize errors. Because some word targets are relatively much more discriminable from the nonword distractors (and vice versa), the subject can set two criteria that will allow rapid decisions for at least some of the stimuli being presented. Thus, a low criterion could be set so that very few words will have FM values that would fall below this criterion. Similarly, a high criterion could be set so that very few nonwords will have FM values that would fall above this criterion. The location and utility of these criteria will be determined by the similarity between the words and nonwords within the test list.

The first stage of the decision process involves a global computation of the FM value of the letter string. That is, the subject makes a quick check to determine if the stimulus is producing any meaning or is very familiar, that is, "Have I seen this stimulus frequently?" If the computed FM value exceeds the upper criterion, the subject will make a fast *word* response; if it fails to exceed the lower criterion, the subject will make a fast *nonword* response. On the other hand, if this FM value falls between the upper and lower criteria, then the subject needs more information before a decision can be made. The necessary information is obtained by performing a more analytic evaluation of the letter string. For example, the subject may actually need to check the spelling of the letter string against the spelling of a word contained in the subject's lexicon. This extra analysis, of course, requires addi-

tional time; thus longer latencies will be found for those words and nonwords requiring such analysis.

Within the present framework, errors can derive from three different sources. First, errors could occur in the global analysis when a word has an extremely low FM value (e.g., *yams*) or when a nonword has an exceptionally high FM value (e.g., *sondpapar*). Second, errors could occur in the analytic stage when there is a lack of knowledge about the appropriate spelling of a word (cf. Gordon, 1983). Third, errors could occur in the analytic stage when the subject has established a criterion time after which a guess will be made because the subject is still unsure about whether the string is spelled correctly or not. Because we feel errors can be produced in all three stages and error rates are generally quite low, we will apply the model primarily to RT data, the data that most currently available models of lexical access have considered to be most important.

The two-stage model we have described provides a straightforward account of the word-frequency effect in the LDT. Low-frequency words have lower values on the FM dimension than high-frequency words. For this reason, the global analysis of a low-frequency word will less often result in a FM value that exceeds the high criterion and permits a rapid *word* response. Thus, a large proportion of the low-frequency words, compared to a smaller proportion of high-frequency words, will require further processing in the analytic stage. The net effect will be that RTs for low-frequency words will be, on average, longer than those for high-frequency words.

The subject's situation in the category verification task is completely different. A discrimination between words and nonwords is not required, so familiarity information is not used in the same manner. In category verification, it is clear that the meaningfulness of a word (concept familiarity) is relevant. That is, a category membership judgment is made about the "meaningfulness" of associating a particular word's concept with a category, but frequency of occurrence in print (word familiarity) is not relevant to the judgment being made. Word frequency could affect the time to determine what is known about a word (lexical access), but this encoding process may be a very small part of the overall judgment process. Thus, there would be only a very small unique LFREQ effect. Similarly, in reading, where the primary task is the extraction of meaning and not the discrimination between words and nonwords, frequency of occurrence is not a meaningful dimension for the task at hand; thus only a very small unique effect of word frequency would be expected.

The present framework has the important feature that it can account for a number of results from the lexical decision literature that have been problematic for one or more of the available models of word recognition. Although it is beyond the scope of the present article to address all of this literature, it is useful to describe briefly how the model accounts for some of the most important findings.

First, within the present framework what should occur if one increases the separation of the word and nonword distributions along the FM dimension by lowering the nonword distribution? This should allow the subject to reduce the upper criterion without increasing the nonword error rate and should have the effect of reducing the proportion of words requiring the slower analytic check process. Because there are more low-frequency words than high-frequency words with FM values below the original upper criterion, the reduced upper criterion will affect more decisions about low-frequency words than those about high-frequency words. The relative reduction in low-frequency word RT should thus be greater than that for high-frequency word RT. One way to lower the nonword distribution would be to present unpronounceable nonwords (*stpne*) as opposed to pronounceable nonwords (*penst*). In support of the present analysis both James (1975) and Duchek and Neely (1984) have found that low-frequency word decisions are facilitated more than are high-frequency word decisions by the presence of unpronounceable nonwords.

A second variable that may influence the subject's placement of the upper and lower criteria is the frequency blocking manipulation (Glanzer & Ehrenreich, 1979; Gordon, 1983). In studies using this technique, performance with pure lists (either only high- or only low-frequency words) is compared with perfor-

mance with mixed lists (both high- and low-frequency words). First, consider the pure lists of high-frequency words. Because the difficult discrimination between low-frequency words and nonwords in a mixed list demands further analysis for many decisions, eliminating the low-frequency words from the list may encourage the subject to relax both criteria, that is, lower the upper criterion and raise the lower criterion. This relaxation of the criteria would reduce latencies both for high-frequency words and for nonwords. Both Glanzer and Ehrenreich (1979) and Gordon (1983) reported data for pure high-frequency lists that match this prediction. For the pure low-frequency lists, on the other hand, subjects cannot shift their criteria because the difficult discrimination between low-frequency words and nonwords remains. Glanzer and Ehrenreich found a small, but nonsignificant, effect of blocking for the low-frequency words, but Gordon found a 0-ms difference between mixed and pure lists for low-frequency words, precisely as the model predicts. Interestingly, Forster (1981) used the pronunciation task that does not involve the decision stage and did not find a blocking effect.[4]

A third important finding in the lexical decision literature has been the impact of repeating some of the words and/or nonwords within the experiment. Repeating a word or nonword should have the effect of increasing its FM value. For words, an increase in the FM value should reduce average RT for low-frequency words more than for high-frequency words because, as indicated earlier, decisions about low-frequency words are more likely to need an analytic check. This is precisely the pattern found by Scarborough, Gerard, and Cortese (1979) and Scarborough, Cortese, and Scarborough (1977). For nonwords, repetition should increase the likelihood that the FM value of a repeated nonword will exceed the lower criterion. When this happens, the repeated nonword latency will be increased because of the need for an analytic check. Duchek and Neely (1984), Durgunoglu (1982), and McKoon and Ratcliff (1979) have all obtained this deleterious effect of repetitions on nonword latencies.[5]

Repetition is not the only way to increase the FM value of a word. For example, contextual priming should increase FM values. Thus, the FM value of the word *cat* should be higher when immediately preceded by *dog* than when preceded by *frog*. Such relatedness effects have been reported in a number of lexical decision experiments (e.g., Balota, 1983; Meyer & Schvaneveldt, 1971; Neely, 1977). Within the present framework, the influence of semantic context should be the greatest for those words that necessitate the more detailed analytic check, primarily the low-frequency words. Evidence in accord with this view has been provided by Becker (1979), who found that contextual effects were significantly larger for low- than for high-frequency words. In a second highly relevant study, Shulman and Davison (1977) investigated contextual priming and nonword pronounceability. As noted previously, one would expect that targets with an unrelated context should be most likely to go through an analytic check. If the upper criterion has been reduced because unpronounceable nonwords are being presented, the major effect of the criterion shift should be on decision times for targets preceded by unrelated words. Shulman and Davison (Experiment 1) reported that the change from pronounceable nonwords to unpronounceable nonwords produced a reduction in latency of 119 ms for the unrelated targets but only 46 ms for the related targets.

---

[4] In contrast to Forster's study, it should be noted that Berry (1971) also used a pronunciation task and found a main effect of blocking and frequency with no evidence of an interaction. Unfortunately, Berry used only 12 words in his study and had pairs of words with the same first letter in the mixed list but not in the pure list. Because the words were presented randomly, it is possible that there was mispriming in the mixed list of a particular pronunciation by these matched pairs (e.g., *above, abet, been, beige, sedate, season*). Moreover, 3 of the 6 low-frequency words were phonologically irregular (*abet, beige, sedate*), and these three items showed the largest blocking effect. In fairness to Berry, the blocking effect was not of primary interest in his study.

[5] Scarborough, Gerard, and Cortese (1979) actually found a slight facilitation for nonword repetitions at short lags. However, these repetition effects disappeared at longer lags. Nonword repetition effects at short lags may be response priming effects. The nonword repetition inhibition effects reported by Duchek and Neely (1984), Durgunoglu (1982), and McKoon and Ratcliff (1979) occur when the nonwords are presented in an earlier list of materials and then later are presented in an LDT.

This brief discussion of some of the lexical decision literature bearing critically on the framework we have proposed is, of course, only the first step in testing the adequacy of the model. However, at this stage, the model does appear to be a useful alternative to other viewpoints because it can provide new insights into findings from lexical decision experiments that have been troublesome for dominant theories of word recognition. The model says nothing about the lexical access process and, in that sense, is relatively uninteresting. Its contribution lies in helping to understand the complexities of the LDT and in evaluating the relevance of results from the LDT to understanding the lexical access process. A full-scale development and quantitative evaluation of the model can be pursued, but the continued importance of the LDT to the study of word recognition and memory will determine the importance of such a pursuit. We believe that the LDT is a useful tool for the study of many kinds of questions, for example, the study of priming effects, and we expect that many researchers agree with us. This being the case, one can expect that the model will be submitted to rigorous empirical test and theoretical analysis.

Conclusions

Three experiments have been reported that investigated the impact of the same five variables for the same set of words across three different tasks. The results indicated that there were striking differences in the unique effect of word frequency across the tasks even though each of the tasks should involve a similar lexical access process. Particularly dramatic was the large unique influence of word frequency in the LDT and its negligible unique effect on the RT for *no* responses in the category verification task. It was argued that the word-frequency effect may be exaggerated in the LDT because of the importance of familiarity of the stimulus in discriminating targets from distractors. A simple two-stage model of the LDT was described that adequately accounts for a number of LDT results that are problematic for most currently dominant models of lexical access.

It is important to reiterate here that we are not arguing that word frequency has no impact on lexical access. Our data do not support that conclusion, and our model takes no position with respect to such a claim. There are data from other tasks that have been taken as evidence that word frequency has a strong impact on lexical access. Two such tasks are tachistoscopic word recognition (Broadbent, 1967; Broadbent & Broadbent, 1975) and pronunciation (Andrews, 1982; Frederiksen & Kroll, 1976). However, there are also problems with the evidence from both of these tasks. First, some researchers (e.g., Catlin, 1969, 1973) believe that the effect of word frequency found in tachistoscopic studies is a response selection effect and not a lexical access effect. Second, the threshold word recognition task provides data relevant to the study of the effect of word frequency on probability of reporting a word from a given frequency class, and these data are not necessarily relevant to "speed" of lexical access, the issue addressed by most current models. Third, there is now mounting evidence (see Balota & Chumbley, in press; Theios & Muise, 1977) suggesting that the word-frequency effect in the pronunciation task stems, at least in part, from the production stage of pronunciation rather than only the lexical access stage.

In sum, the major thrust of the present article has been to point out the importance of considering the decision task the subject faces in making lexical decisions. Unless one considers the decision stage, the results from the LDT may misdirect the development of an adequate model of lexical access. In this light, the task confronting those interested in word recognition is either to (a) demonstrate the relevance of lexical decision results to lexical access while taking into consideration the decision stage or (b) develop a different task that more faithfully reflects the processes involved in word recognition.

References

Anderson, J. R., & Reder, L. M. (1974). Negative judgments in and about semantic memory. *Journal of Verbal Learning and Verbal Behavior, 13,* 664–681.

Andrews, S. (1982). Phonological recoding: Is the regularity effect consistent? *Memory & Cognition, 10,* 565–575.

Atkinson, R. C., & Juola, J. F. (1973). Factors influencing

speed and accuracy of word recognition. In S. Kornblum (Ed.), *Attention and performance IV* (pp. 583–612). New York: Academic Press.

Balota, D. A. (1983). Automatic semantic activation and episodic memory encoding. *Journal of Verbal Learning and Verbal Behavior, 22,* 88–104.

Balota, D. A., & Chumbley, J. I. (in press). The locus of word-frequency effects in the pronunciation task: Lexical access and/or production frequency? *Journal of Verbal Learning and Verbal Behavior.*

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph, 80,*(3, Pt. 2).

Becker, C. A. (1976). Allocation of attention during visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 556–566.

Becker, C. A. (1979). Semantic context and word frequency effects in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 252–259.

Becker, C. A. (1980). Semantic context effects in visual word recognition: An analysis of semantic strategies. *Memory & Cognition, 8,* 493–512.

Becker, C. A., & Killion, T. H. (1977). Interaction of visual and cognitive effects in word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 389–401.

Berry, C. (1971). Advanced frequency information and verbal response times. *Psychonomic Science, 21,* 151–152.

Broadbent, D. E. (1967). Word-frequency effect and response bias. *Psychological Review, 74,* 1–15.

Broadbent, D. E., & Broadbent, M. H. P. (1975). Some further data concerning the word-frequency effect. *Journal of Experimental of Psychology: General, 104,* 297–308.

Carpenter, P. A., & Just, M. A. (1983). What your eyes do while your mind is reading. In K. Rayner (Ed.), *Eye movements in reading: Perceptual and language processes* (pp. 275–307). New York: Academic Press.

Catlin, J. (1969). On the word-frequency effect. *Psychological Review, 76,* 504–506.

Catlin, J. (1973). In defense of sophisticated-guessing theory. *Psychological Review, 80,* 412–416.

Chumbley, J. I. (1984). *The roles of typicality, instance dominance, and category dominance in verifying category membership.* Manuscript in preparation.

Chumbley, J. I., & Balota, D. A. (1984). *A word's meaning affects the decision in lexical decision.* Manuscript submitted for publication.

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535–555). New York: Academic Press.

Duchek, J. M., & Neely, J. H. (1984). *Levels-of-processing and frequency effects in episodic and semantic memory.* Manuscript in preparation.

Durgunoglu, A. (1982). *Episodic and semantic priming in episodic and semantic memory.* Unpublished master's thesis. Purdue University.

Forster, K. I. (1976). Accessing the mental lexicon. In R. J. Wales & E. Walker (Eds.), *New approaches to language mechanisms* (pp. 257–287). Amsterdam: North-Holland.

Forster, K. I. (1979). Levels of processing and the structure of the language processor. In W. E. Cooper & E. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 27–85). Hillsdale, NJ: Erlbaum.

Forster, K. I. (1981). Frequency blocking and lexical access: One mental lexicon or two? *Journal of Verbal Learning and Verbal Behavior, 20,* 190–203.

Frederiksen, J. R., & Kroll, J. F. (1976). Spelling and sound: Approaches to the internal lexicon. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 361–379.

Glanzer, M., & Ehrenreich, S. L. (1979). Structure and search of the internal lexicon. *Journal of Verbal Learning and Verbal Behavior, 18,* 381–398.

Gordon, B. (1983). Lexical access and lexical decision: Mechanisms of frequency sensitivity. *Journal of Verbal Learning and Verbal Behavior, 22,* 24–44.

James, C. T. (1975). The role of semantic information in lexical decisions. *Journal of Experimental Psychology: Human Perception and Performance, 1,* 130–136.

Kliegl, R., Olson, R. K., & Davidson, B. J. (1982). Regression analyses as a tool for studying reading processes: Comment on Just and Carpenter's eye fixation theory. *Memory & Cognition, 10,* 287–296.

Kliegl, R., Olson, R. K., & Davidson, B. J. (1983). On problems of unconfounding perceptual and language processes. In K. Rayner (Ed.), *Eye movements in reading: Perceptual and language processes* (pp. 333–343). New York: Academic Press.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

Loftus, E. F., & Suppes, P. (1972). Structural variables that determine the speed of retrieving words from long-term memory. *Journal of Verbal Learning and Verbal Behavior, 11,* 770–777.

McKoon, G., & Ratcliff, R. (1979). Priming in episodic and semantic memory. *Journal of Verbal Learning and Verbal Behavior, 18,* 463–480.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90,* 227–234.

Millward, R. B., Rice, G., & Corbett, A. (1975). Category production measures and verification times. In R. A. Kennedy and R. W. Wilkes (Eds.), *Studies in long-term memory* (pp. 219–252). New York: Wiley.

Mitchell, D. C., & Green, D. W. (1978). The effects of context and content on immediate processing in reading. *Quarterly Journal of Experimental Psychology, 30,* 609–636.

Morton, J. (1969). The interaction of information in word recognition. *Psychological Review, 76,* 165–178.

Morton, J. (1970). A functional model for memory. In D. A. Norman (Ed.), *Models of human memory* (pp. 203–254). New York: Academic Press.

Morton, J. (1982). Disintegrating the lexicon: An information processing approach. In J. Mehler, S. Franck, E. C. T. Walker, & M. Garrett (Eds.), *Perspectives on*

mental representation (pp. 89–109). Hillsdale, NJ: Erlbaum.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited capacity attention. *Journal of Experimental Psychology: General, 106,* 226–254.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104,* 192–203.

Sanford, A. J., & Seymour, P. H. K. (1974). Semantic distance effects in naming superordinates. *Memory & Cognition, 2,* 714–720.

Scarborough, D. L., Cortese, C., & Scarborough, H. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 1–17.

Scarborough, D. L., Gerard, L., & Cortese, C. (1979). Accessing lexical memory: The transfer of word repetition effects across task and modality. *Memory & Cognition, 7,* 3–12.

Shulman, H. G., & Davison, T. C. B. (1977). Control properties of semantic coding in a lexical decision task. *Journal of Verbal Learning and Verbal Behavior, 16,* 91–98.

Theios, J., & Muise, J. G. (1977). The word identification process in reading. In N. J. Castellan, D. B. Pisoni, & G. R. Potts (Eds.), *Cognitive theory.* (Vol. 2, pp. 289–327). Hillsdale, NJ: Erlbaum.

West, R. F., & Stanovich, K. E. (1982). Source of inhibition in experiments on the effect of sentence context on word recognition. *Journal of Experimental Psychology: Human Learning, Memory, and Cognition, 8,* 385–399.

Whaley, C. P. (1978). Word-nonword classification time. *Journal of Verbal Learning and Verbal Behavior, 17,* 143–154.

Received July 29, 1983
Revision received December 5, 1983 ■

Instructions to Authors

Authors should prepare manuscripts according to the *Publication Manual of the American Psychological Association* (3rd ed.). All manuscripts must include an abstract of 100–150 words typed on a separate sheet of paper. Typing instructions (all copy must be double-spaced) and instructions on preparing tables, figures, references, metrics, and abstracts appear in the *Manual.* Also, all manuscripts are subject to editing for sexist language. For further information on content, authors should refer to the editorial in the February 1982 issue of the *Journal* (Vol. 8, No. 1, pp. 171–172). For information on the other three JEP journals, authors should refer to editorials in those journals.

APA policy prohibits an author from submitting the same manuscript for concurrent consideration by two or more journals. APA policy also prohibits duplicate publication, that is, publication of a manuscript that has already been published in whole or in substantial part in another publication. Authors of manuscripts submitted to APA journals are expected to have available their raw data throughout the editorial review process and for at least 5 years after the date of publication.

Blind reviews are optional, and authors who wish blind reviews must specifically request them when submitting their manuscripts. Each copy of a manuscript to be blind reviewed should include a separate title page with authors' names and affiliations, and these should not appear anywhere else on the manuscript. Footnotes that identify the authors should be typed on a separate page. Authors should make every effort to see that the manuscript itself contains no clues to their identities.

Manuscripts should be submitted in quintuplicate (the original and four photocopies), and all copies should be clear, readable, and on paper of good quality. Authors should keep a copy of the manuscript to guard against loss. Mail manuscripts to the Editor, William Epstein, Department of Psychology, University of Wisconsin, W. J. Brogden Psychology Building, 1202 West Johnson Street, Madison, Wisconsin 53706.

For the other JEP journals, authors should submit manuscripts (in quadruplicate) to one of the editors at the following addresses: *Journal of Experimental Psychology: General,* Sam Glucksberg, Department of Psychology, Princeton University, Princeton, New Jersey 08544; *Journal of Experimental Psychology: Learning, Memory, and Cognition,* Richard M. Shiffrin, Department of Psychology, Indiana University, Bloomington, Indiana 47405; and *Journal of Experimental Psychology: Animal Behavior Processes,* Donald S. Blough, Department of Psychology, Brown University, Providence, Rhode Island 02912. When one of the editors believes a manuscript is clearly more appropriate for an alternative journal of the American Psychological Association, the editor may redirect the manuscript with the approval of the author.